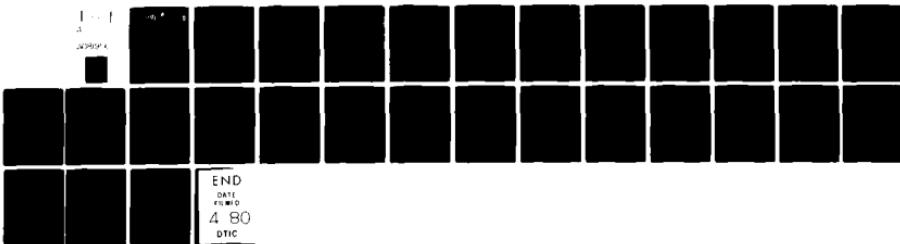


AD-A081 956

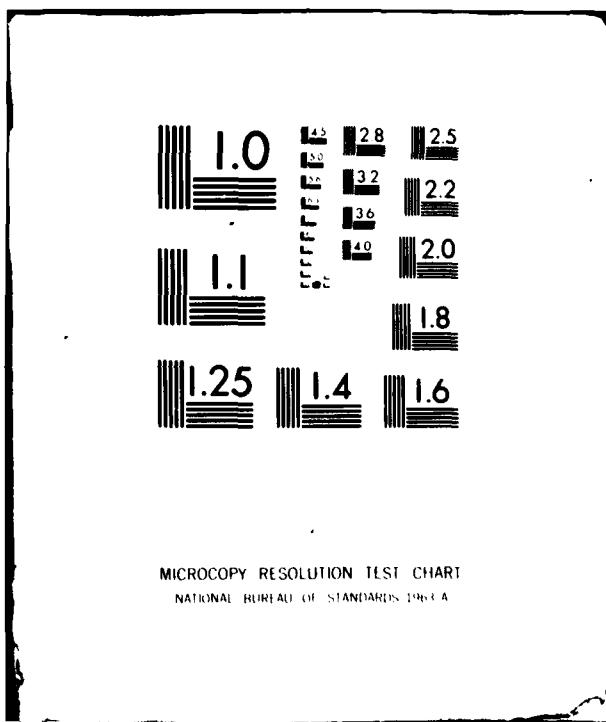
ILLINOIS UNIV AT URBANA-CHAMPAIGN COORDINATED SCIENCE LAB F/8 5/8
VISUAL ANALOG REPRESENTATIONS FOR NATURAL LANGUAGE UNDERSTANDING--ETC(U)
FEB 79 D L WALTZ, L BOGESS N00014-75-C-0612

UNCLASSIFIED WP-20 NL

1
A
200004



END
DATE
4 80
DTIC



ADA081956

LEVEL

⑨ WORKING PAPER 20

⑥ Visual Analog Representations for
Natural Language Understanding

⑩ David L. Waltz and Lois Boggess

Advanced Automation Group
Coordinated Science Laboratory,
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

⑪ February 1979

⑭ WP-20

⑫ 29

Abstract

In order for a natural language system to truly "know what it is talking about," it must have a connection to the real-world correlates of language. For language describing physical objects and their relations in a scene, a visual analog representation of the scene can provide a useful target structure to be shared by a language understanding system and a computer vision system.

This paper discusses the generation of visual analog representations from input English sentences. It also describes the operation of a LISP program which generates such a representation from simple English sentences describing a scene. A sequence of sentences can result in a fairly elaborate model. The program can then answer questions about relationships between the objects, even though the relationships in question may not have been explicit in the original scene description. Results suggest that the direct testing of visual analog representations may be an important way to bypass long chains of reasoning and to thus avoid the combinational problems inherent in such reasoning methods.

Key Words and Phrases:

Natural language understanding, language and perception, scene description, representation of knowledge, frames.

*This work is supported by the Office of Naval Research under Contract N00014-75-C-0612

†Current address: Dr. Lois Boggess, Computer Science Department, Mississippi State University, Mississippi State, Mississippi 39762

Working papers are informal papers intended for internal use.

This document has been approved
for public release and is free
of classification or restriction

091700

1 03 12 040

80 3 7 085

1. Understanding language about the physical world

Suppose that we are given a sentence such as (1):

(1) A dog bit a mailman.

How do we understand such a sentence? What inferences do we make and what inferences can we make? To help answer these questions, suppose that we are asked:

(2) Where on his body did the dog probably bite the mailman?

We suggest that most people would answer (2) with: "on the leg," and that as a first guess, such an answer could plausibly be part of a BITING script or default slots of a frame for BITING. However, suppose that we insert one or more of the following sentences after (1) before asking (2):

(3 a,b) The dog was a {doberman.
dachshund.}

(3 c,d,e) The man was {sitting.
lying down.
3 feet tall.}

(3 f,g) The dog was {standing on its hind legs.
sitting} at the time.

In these cases the answers to (2) could be quite different parts of the body ("arm" becomes most likely if bitten by a doberman) or one could be much more definite about the answer ("leg" becomes overwhelmingly likely if one is bitten by a dachshund while standing up).

How could we successfully model in a program the understanding process a person goes through in this example? We suggest that the simplest and most natural way to model this understanding is to build up a visual analog knowledge base (representing "person," "dog," etc. as 3-D spatial entities) and to write programs which can manipulate and integrate these visual analog representations. For the example given in (1) - (3), figure 1 illustrates some of the information that would have to be included.

Figure 1

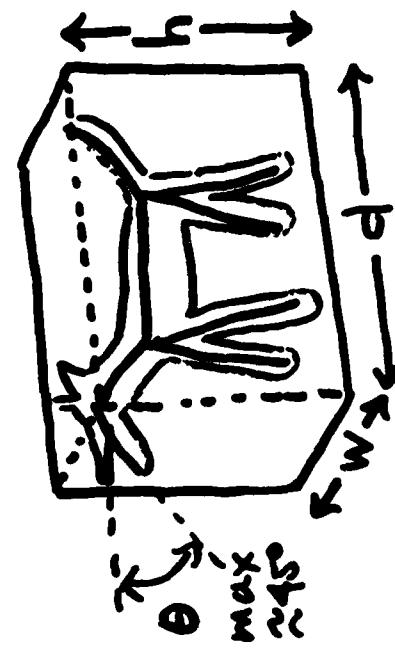
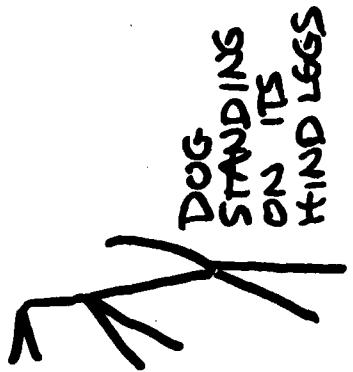
+ knowledge about r_t , r_a , r_s , etc.

+ bite (x, y) means

put (x , around (mouth(x), part-of(y)))
& apply-force (x , mouth-joint(x))

+ "x feet tall" $\rightarrow h = x$

+ ...



DACHSHUND : 3
DOBERMAN : 4

$\frac{h}{w}$

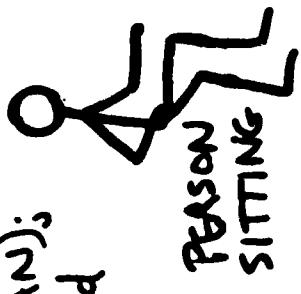
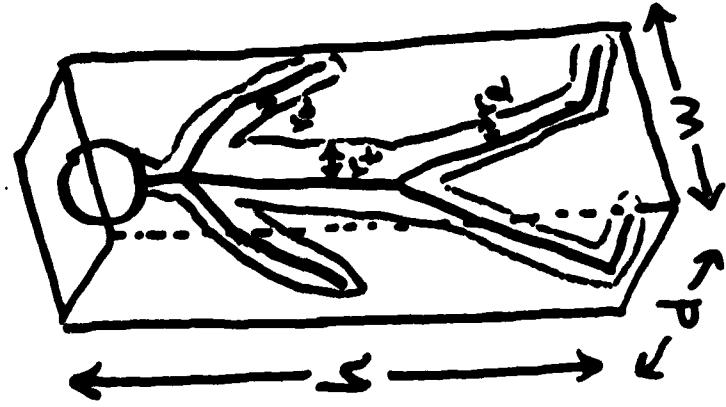
$\frac{1}{1}$

$\frac{1}{3}$

$\frac{1}{4}$

$\frac{1}{3}$

$\frac{1}{3}$



As another example, suppose we were given the following set of sentences:

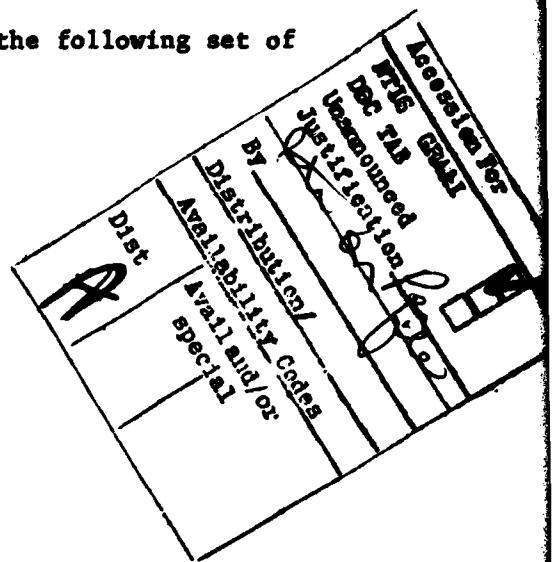
- (4) A goldfish is in a goldfish bowl.
- (5) The goldfish bowl is on a shelf.
- (6) The shelf is on a desk.
- (7) The desk is in a room.

Now suppose that we are asked:

- (8) Is the goldfish in the room? or
- (9) Is the goldfish on the desk?

The answer to (8) should of course be "yes" and the answer to (9) should be something like "Not directly on, but on is still an appropriate description." How could we mechanize the answering of such questions?

We suggest that these questions can be easily answered if we use (4) - (7) to build an integrated visual analog structure which represents (4) - (7). Then given procedural definitions of prepositions like in and on, with the "subject" and "object"** that the prepositions relate viewed as arguments to the procedures, we can apply the procedures to the structure directly to answer the questions. Boggess [1978] has written a program, described later in this paper, which works exactly in this manner. Given (4) - (7), the program constructs an analog model like that shown in figure 2.



*Given the phrase "the cat on the mat," or the sentence, "The cat is on the mat," we call cat the subject and mat the object.

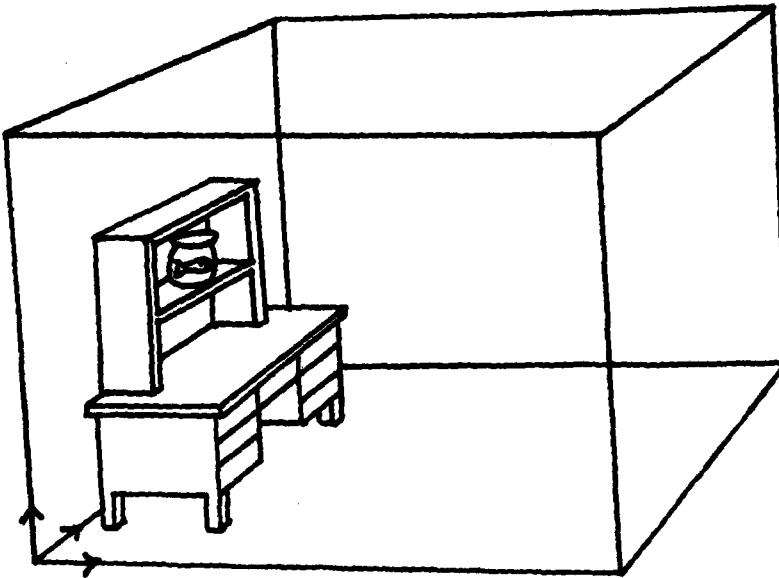


Figure 2

In constructing the model, the program uses properties of the subject and object to decide what the preposition means in each given case. For example, on would be interpreted quite differently in "the shelf on the wall," and "the shadow on the wall." "The shelf" is assumed to touch the wall, be supported by it, and to have a preferred orientation and position (height) with respect to the wall, whereas none of these are true of "shadow." The differences in treatment are the result of noting in the lexicon that "shelf" is an ordinary physical object (requiring support) with preferred orientation--its free surface should be horizontal--whereas a "shadow" is 2-D and weightless and thus does not require support. Each object has default dimensions as well as weight, and these dimensions (actually a rectangular parallelepiped which encloses the object) can be used to construct the visual analog model.

To illustrate the interpretation of this model, consider the preposition in. If in's object is a 3-D enclosure, then all we need to do to

see whether the subject is in the object is to check whether the coordinates of all the corners of the subject are within the intervals of the coordinates of the corners of the object. The answer can be found with one set of tests, regardless of how many chained statements were required to relate the subject and object in the scene description. Thus, given a model like figure 2, it is very easy to answer (8) because all the dimensions of "goldfish" are within the dimensions of "room."

2. Using deductive rules on a data base of assertions.

Starting with Black [1968], there have been programs which dealt with similar questions. Most of these programs have "understood" sentences like (4) - (7) by adding something equivalent to an assertion of the form (ON GOLDFISH-BOWL1 SHELF1) to a data base. Answering questions about the scene described has then involved applying deductions rules such as:

$$(10) \text{ (ON ?A ?B) AND (ON ?B ?C)} \Rightarrow \text{(ON ?A ?C)}$$

to verify that a given relationship does or does not hold between two given objects. In general, the set of assertions in the data base will define a network, i.e. any two items in the data base may be connected by an arbitrary number of deductive chains or direct assertions. For example, a chair can at the same time be at a desk, under the desk and touching the desk. There are at least two serious difficulties with using a method like deductive chaining to understand the spatial domain, represented as a data base of assertions:

A. If there are many rules and many objects, the search for a deductive chain which can prove or disprove a given relation between two objects can involve combinational explosion. Often there will be insufficient information to decide whether a relationship holds between two objects;

in such a case, all relevant paths between the objects will have to be explored before a system can decide that the problem cannot be decided.

B. Even more serious is the difficulty in formulating deduction rules properly to begin with. For example, rule (10) allows us to deduce correctly that a leaf is on a tree if the leaf is on a branch and the branch is on a tree, but it is not correct to deduce that a cow has wings if we know that a wing is on a fly and the fly is on a cow!

One obvious solution to this difficulty has been to create a number of definitions for ON--ON1, ON2, ON3, and so on, where ON1 might mean "is a part of" as in "the wing on a fly," ON2 might mean "above, touching and supported by" as in "the pencil on the desk," etc. Deduction rules can then be formulated with greater precision, but we have added an additional problem: when on is asserted to hold between two objects or used in a question a program must now decide whether ON1, ON2, ON3, or ONn is intended. More rules delimiting the classes of objects which can be related by each meaning of on then have to be formulated and somehow utilized to decide which meaning(s) are appropriate.

But even a large number of such rules cannot easily substitute for the visual analog model. Suppose that (4)-(6) were followed by

(11) The desk is in a box.

In this case the goldfish may or may not be inside the box, depending on the dimensions of the box, desk, and shelf (see figure 3).

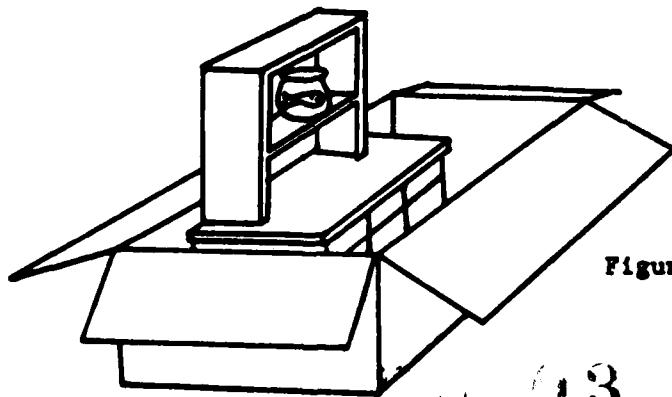


Figure 3

13 12 040

But how could a deduction-rule-based system give a different answer to these two cases, unless it implicitly coded metric information? And if it coded metric information, why bother with the potentially long deductive chains?

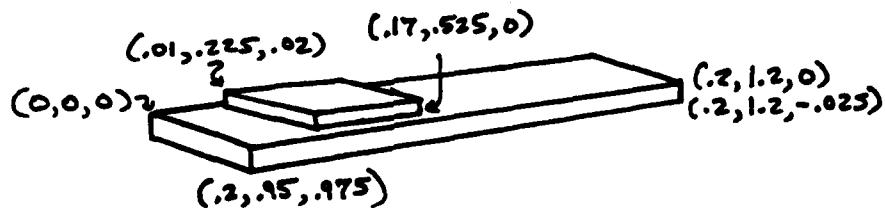
3. Operation of a Program for understanding simple language about space.

A MACLISP program has been written by Boggess [1978] which can build a spatial model of sentences involving in and on relations, and answer questions about its model. Input to the program consists of normal English sentences, which are parsed with the aid of a LINGOL [Pratt 1973] preprocessor. LINGOL is an MIT-originated program package which accepts grammatical rules of the type $S \rightarrow NP + VP$ and produces LISP programs which can then parse input sentences according to the rules of the specified grammar. For this implementation, LINGOL was used to single out prepositions and their semantic subject and object. For example, in the sentence, "On the bed was a box," the LINGOL portion passes the preposition on, the semantic object the-bed and the semantic subject the-box to the rest of the program.

Some examples

Suppose a user types: A book is on the shelf.

As the result of this input, an individual book and individual shelf are created; the modeling portion of the program records the location restrictions for the book, chooses a location for it and gives the user the global coordinates of the book and shelf. These correspond to the following illustration:

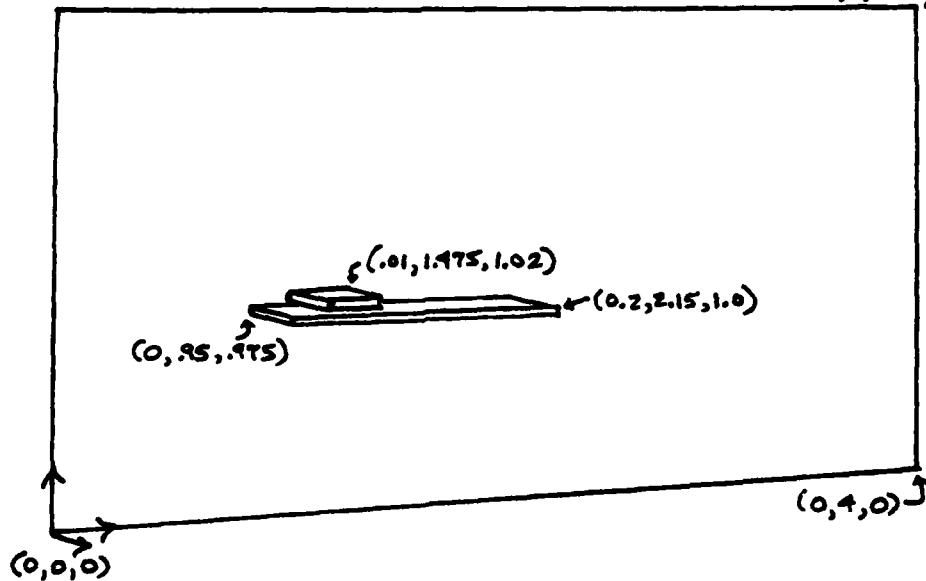


Notice the upper surface of the shelf is chosen for the $Z = 0$ plane--leaving all but the top surface of the shelf below the plane with negative coordinates. It is entirely possible that we are about to enter an extended description of many objects, all of which are on the shelf or above it, in which case treating the top surface of the shelf as our basic horizontal plane makes sense.

Suppose now that the next sentence is: The shelf is on a wall.

Since the object that has been serving as our origin has just been treated as a semantic subject and related to another object, the location of the shelf and everything related to it is accordingly revised. We use the symbol \nearrow to indicate the origin of the global coordinate system in the illustrations.

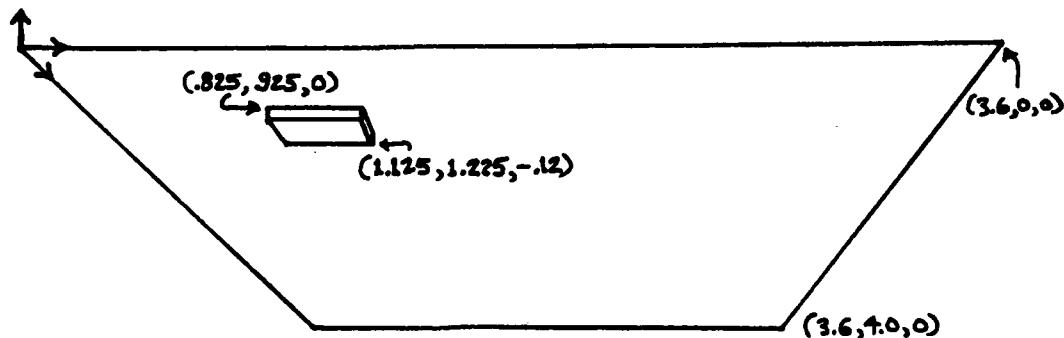
$(0, 4, 2.5) \nearrow$



Actually, in the current implementation, it is merely an accident that the long edge of the shelf is aligned with the wall. But for the order in which the dimensions of the shelf were given in the data, the program might just as well have set the short end of the shelf flush with the wall. However, the program would not place the top surface of the shelf against

the wall, even if it were bare, since that surface is marked as characteristically horizontal. Incidentally, were it not for the fact that shelves are marked as having a characteristic height (a little bit of "world knowledge") the program would have put the shelf considerably lower.

Suppose an input were: A light is on a ceiling.

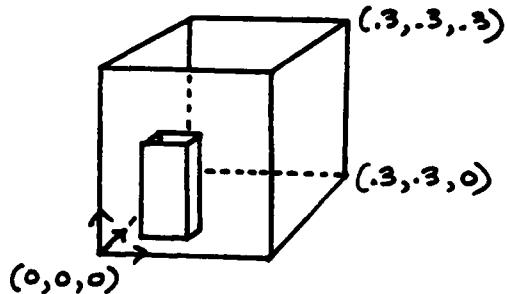


A corner of the ceiling is taken as the origin. Since the ceiling has a marked free-direction vertically downward, the light ends up on the correct side of the ceiling surface. Notice that, while people would ordinarily put the light in the middle of the ceiling, the program doesn't know enough about ceilings and lights to do so.

Finally, let's follow an extended example.

Input: A glass is in a box.

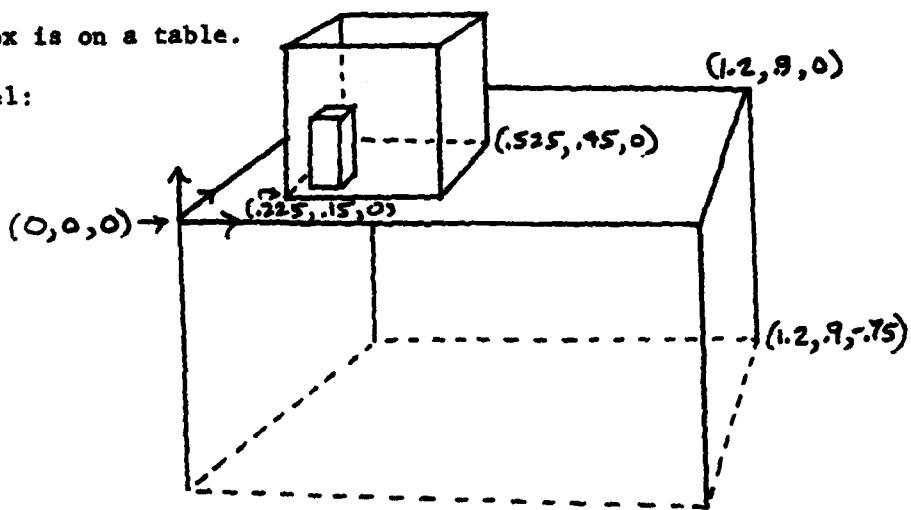
Resulting model:



Comments: The glass has weight, so it ends up not only in the box, but at the bottom of it.

Input: The box is on a table.

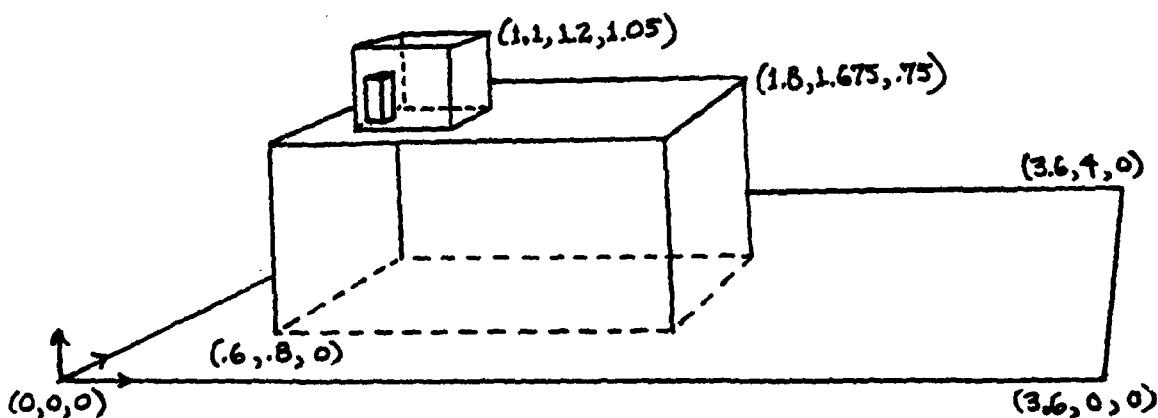
Resulting model:



Comments: There is only one individual box known to the system, so the phrase "the box" can be interpreted with no difficulty. Notice the surface of the table is taken as the basic plane for the discussion so far, rather than putting the origin at, say, a point at the bottom of the table.

Input: The table is on a floor.

Resulting model:



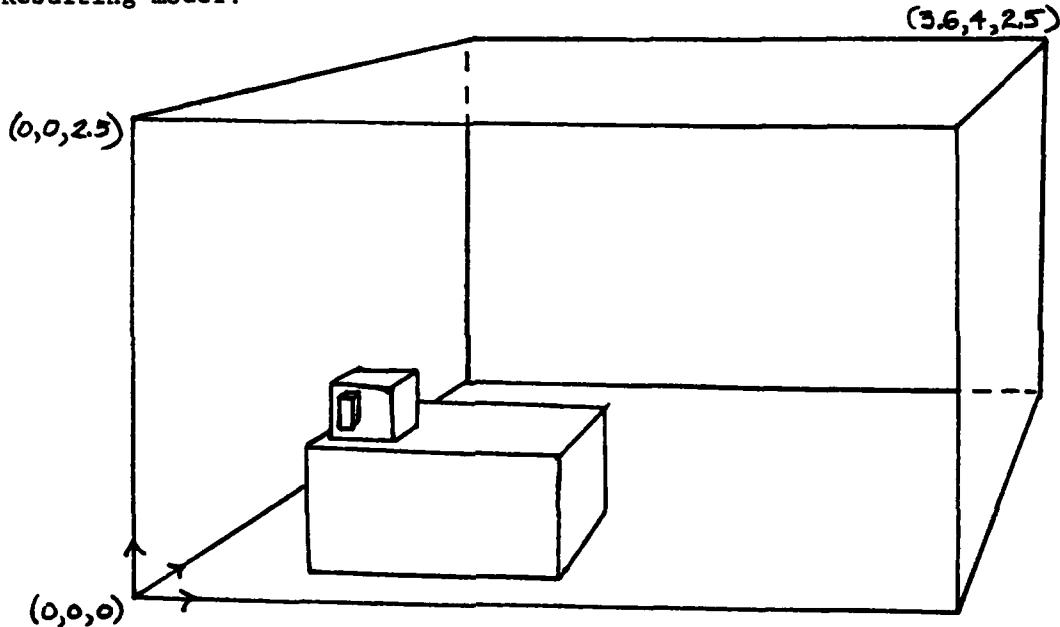
Comments: "A floor" sounds strange, but the system doesn't know for the present that tables are almost always on floors, so mentioning a particular

table does not allow it to presuppose a particular floor that it could reference as the floor.

As is probably becoming obvious, the model does not choose locations randomly. Rather, it tends toward a particular corner. This choice was made in hopes of avoiding the "findspace" problem [Sussman 1973] when several objects must be located on one surface.

Input: The floor is in a room.

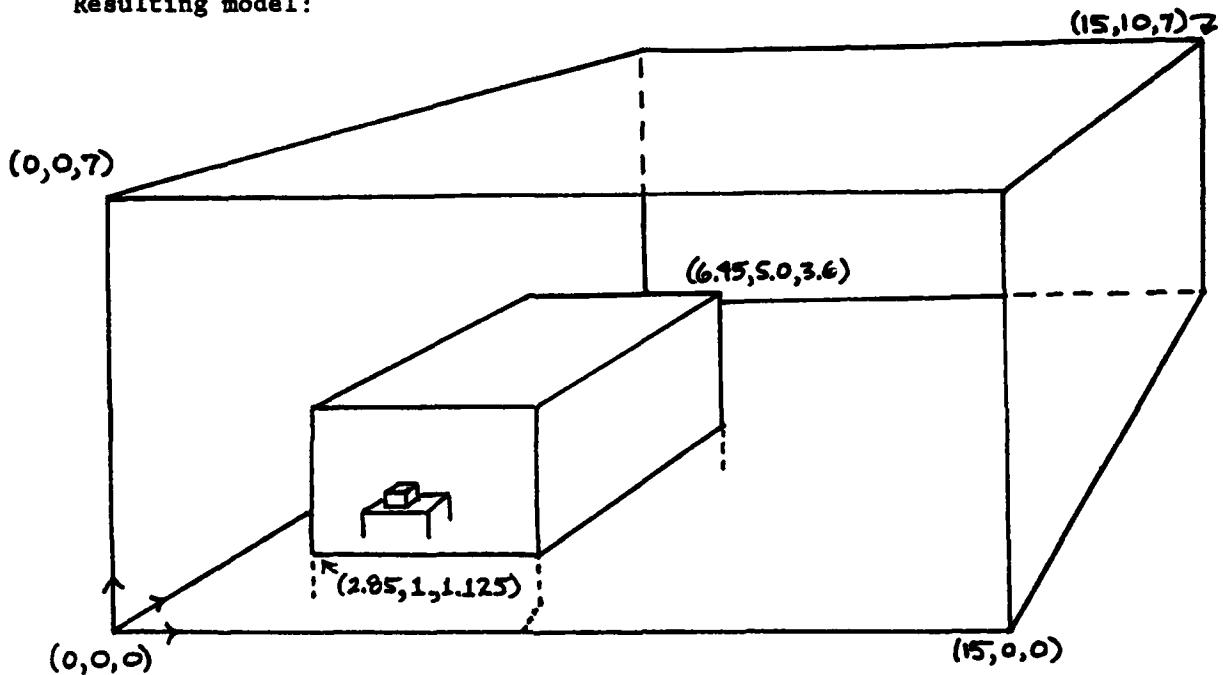
Resulting model:



Comments: Again, the model doesn't know that a floor is part of a room. Naturally, a default-sized floor exactly fits a default-sized room, but the model has to know that a floor belongs at "ground level" or it would try to put the floor at a more or less arbitrary level in the room. While this particular sentence sounds unusual, it is natural to speak, say, of "the floor in Jonathan's room."

Input: The room is in a house.

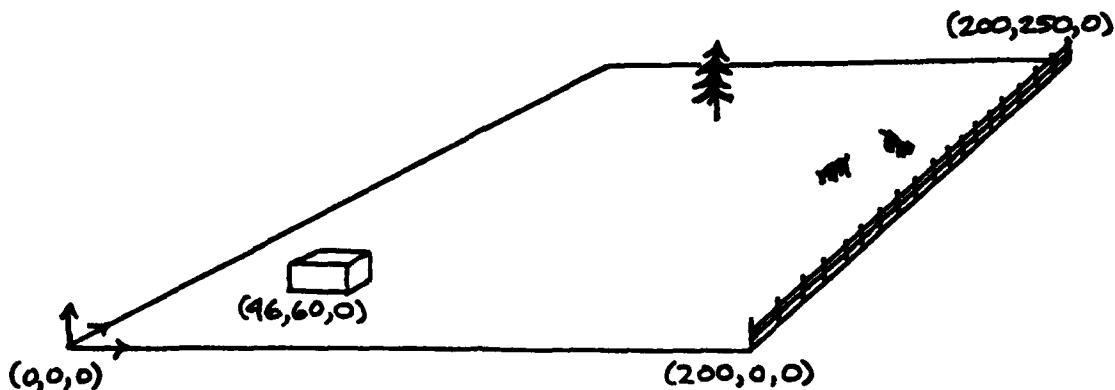
Resulting model:



Comments: It is not very evident from the illustration, but the room is actually several feet off the ground in the model. Obviously, this "goof" could have been fixed (there are several natural solutions that would not have involved extending the capabilities of the model), but it was left in for two reasons--it illustrates what the model does with an in relation involving a weightless object, and secondly, if the building had been a hotel rather than a house, a room in the same relative position in the larger building would have seemed quite reasonable.

Input: The house is in a field.

Resulting Model:



Comments: A house has no weight, either, as far as the model is concerned, but a field is a two-dimensional object, and the in relation implies contiguity under those circumstances.

Questions

Suppose after all this the user types: Is the box on the table?

The response from the system is YES.

To the input: Is the box on the floor? The system responds NO.

Is the box in the room? YES

Is the glass on the table? NOT DIRECTLY, BUT ON IS STILL AN

ACCEPTABLE DESCRIPTION.

The program answers these questions by directly interrogating the three-dimensional model, not by knowing that, say, if A is on B and B is in C then A is probably in C. At no time did we say that the box was in the room. But thanks to the sizes of boxes and tables and the locations

of floors relative to the rest of a room, there is no question but that the box must be in the room in the most rigorous sense of the word.

It is also possible to handle situations which would be difficult for systems based on chained inference rules. For example, this program can distinguish between a glass on a tall object in a box and a glass on a small object in the box. If the tall object were large enough that the glass was exterior to the box, then this sort of model could reasonably balk at calling the glass in the box--or at least hedge, as a person might. A system built on the sort of inference rules mentioned above could have trouble distinguishing between these cases.

4. Program implementation-representation of prepositions and objects

The examples given in the preceding section were from a session with with a small program, written in MACLISP and run on the DEC-10 system at the Coordinated Science Lab. The program consists of about 45 functions, most of them fairly short. Data for the implementation consisted of twenty-two "definitions" of objects and the definitions of the prepositions themselves. Input to the program consists of English sentences--either statements or questions. Statements are expected to be either "naming" statements ("Tweety is a bird" or "Volume-1 is a book") or locative statements ("A book is on a table," "In the room is a bed"). Output is either a set of coordinates for each object in the "mental model" or a response to the question.

Object definitions

Some sample definitions of visually perceptible objects follow (Units of measurement are meters and kilograms):

```
(TABLE PROTOTYPE
  (INSTANCE-OF FURNITURE)
  (CHARACTERISTIC-SHAPE ((HEIGHT 0.75)
    (CROSS-SEC 1.2 0.9)))
  (FREE-SURFACE (((PLANE HORIZONTAL)
    (FREE-DIRECTION /+Z)
    (HEIGHT 0.75)
    (DIMENSIONS 1/2 0.9))))
  (WEIGHT 25.0))

(BOX PROTOTYPE
  (CHARACTERISTIC-SHAPE ((HEIGHT 0.3)
    (CROSS-SEC 0.3 0.3)))
  (FEATURES (CONTAINER OPEN-TOP))
  (WEIGHT 1.0))

(FLY PROTOTYPE
  (CHARACTERISTIC-SHAPE ((HEIGHT 3.0E-3)
    (CROSS-SEC 3.0E-3 5.0E-3)))

(WALL PROTOTYPE
  (CHARACTERISTIC-SHAPE ((HEIGHT 2.5) (CROSS-SEC 4.)))
  (FREE-SURFACE (((PLANE VERTICAL)
    (HEIGHT 2.5)
    (WIDTH 4.)))))

(CEILING PROTOTYPE
  (CHARACTERISTIC-SHAPE ((CROSS-SEC 3.6 4.)))
  (FREE-SURFACE (((PLANE HORIZONTAL)
    (FREE-DIRECTION /-Z)
    (DIMENSIONS 3/6 4.0))))
  (FEATURES ((CHARACTERISTIC-HEIGHT 2.5)))))

(SHELF PROTOTYPE
  (INSTANCE-OF FURNITURE)
  (FEATURES ((CHARACTERISTIC-HEIGHT 1.0)))
  (WEIGHT 1.5)
  (CHARACTERISTIC-SHAPE ((HEIGHT 0.025)
    (CROSS-SEC 0.2 1.2)))
  (FREE-SURFACE (((PLANE HORIZONTAL)
    (HEIGHT 0.025)
    (DIMENSIONS 0.2 1.2)
    (FREE-DIRECTION /+Z)))))
```

At initialization, the components of the characteristic-shapes are used to create a simple "mental picture" of the object, in the form of coordinates of an enclosing right parallelepiped. The coordinates are always given in a particular order: bottom front right, bottom front left, bottom back left, and so on. This permanent mental picture is kept under a "local coordinates" property, with the bottom right front taken as local origin.

The definitions specifically single out planar free surfaces on a free-surface list, since it is impossible to judge from the representation whether a planar surface is a characteristic of the object itself.

Also included in the definitions is an indication of whether the object is essentially hollow as opposed to essentially "solid" throughout. The surfaces of the latter are the boundaries of matter; the surfaces of the former enclose space. The feature CONTAINER is used to indicate an object whose interior is canonically empty. Another feature applies to CONTAINERs only and is used to indicate whether they are OPEN-TOPPED or not.

CHARACTERISTIC-HEIGHT as part of a feature list indicates that an object normally would be found at a given height above the default ground-level (either the floor or the actual ground). Otherwise a clock placed randomly on a wall might end up very close to the floor. After using the program for a while, it became obvious that we needed to have such default characteristic heights for a number of items--clocks, windows, shelves, counters, cabinets, and so forth.

Preposition definitions

Each preposition is defined as a LISP function with the subject and object as arguments. The LISP functions are based on the results of an

extensive analysis of about 20 spatial locative prepositions (see [Bogges 1978]). In this analysis, a number of primitives were identified, such as CONTIGuous, SUPPORTed, INTERIOR (2-D and 3-D), CROSS-SECTION (of objects), PROJECTION (of CROSS-SECTIONS), TRAJECTORY, UP/DOWN, HORIZONTAL, VERTICAL, and various coordinate systems. These primitives (which unfortunately would require far too much space to treat rigorously here) constitute a major result of this research. They will allow us to express neatly the meanings of the approximately 20 locative prepositions analyzed but not yet programmed, and seem on preliminary analysis to be an adequate set for the spatial use of most of the rest of the prepositions as well (prepositions form a closed set).

Each preposition seems to have a default interpretation if its subject and object are unknown, as in "the thingamajig on the whatchamacallit." The default interpretation represents a "pure" case of the prepositional relation--however the preposition can be used to describe a range of physical situations which vary from the "pure" instance by having one or more components of the default case missing or modified. For example, the pure case of above is that in which the SUBJECT is INTERIOR (3-D) but not CONTIGuous to the bottom of a volume defined by projecting the HORIZONTAL CROSS-SECTION of the OBJECT upward VERTICALLY in space for a distance of on the order of 3 times the object's diameter. However, above can also be used to describe a variety of "impure" relationships in a scene, including cases where the subject is merely at a higher level than the object (as in "there are clouds above us") and cases where the 2-D projected image of the SUBJECT is INTERIOR (2-D) to the region defined by projecting the HORIZONTAL extreme of the 2-D projection of the OBJECT upward VERTICALLY (as in "the moon above Miami").

To give a better idea of what each prepositional definition is like, let us look at what the functions for on and in do. On is faced with two decisions: it must decide which surface of the object the subject is contiguous to, and it must decide which side of the subject is contiguous to the object.

If the subject does not behave normally with respect to gravity (shadows, visual patterns, thin films of liquids and many insects exhibit gravity-defying behavior) then any available surface of the object will do.

If the subject is under gravitational constraints, then the routine looks for one of four possibilities: in order of preference, 1) a horizontal plane in the object, 2) if the object is three-dimensional and is not an open-topped container, then the top of the object 3) failing either of these, then any planar free-surface, and finally 4) any available surface. In any of these cases, the object requires support and by supposition the semantic object furnishes it.

Having found the surface of the object, on looks for a probable surface of the subject. The check to see if the subject has a marked free-surface is actually a back-handed way to see if the subject has a preferred orientation. If it has, the preferred orientation is presumed to be the canonical one, and on passes to a function called CONTIG, not a surface of the subject but the entire subject, thereby instructing CONTIG to translate the subject in whatever direction necessary to bring it into contact with the object-surface indicated, but not to rotate it in any way. On the other hand, if the subject has no preferred orientation, on selects the canonical bottom of the subject.

The definition of the preposition in has to decide if it is dealing with a container, whether the container is open-topped, and whether the

subject behaves normally with respect to gravitational constraints. It then calls one of the INTERIOR functions and, sometimes, CONTIG (when the subject is assumed to be in the bottom of a container, for instance). At present, the system has two- and three-dimensional interior functions, which restrict the location of their subject with respect to a plane of their object or the volume delineated by the object, respectively.

5. Assessment of the program.

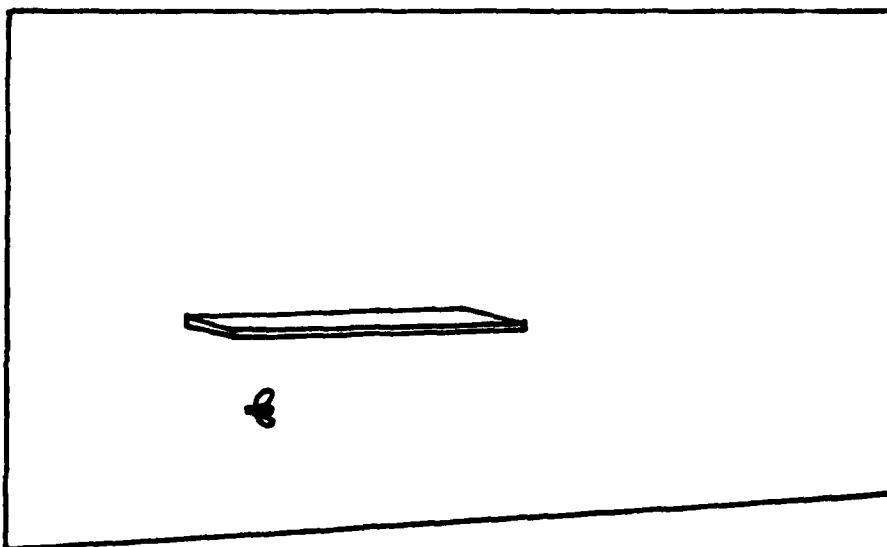
Inferencing problems

One of the nice features of this "analog model" is that it holds out hope for doing inferencing and deduction by direct reference to the model, under optimum conditions, and by reference to the model plus the location restrictions under other circumstances; the construction of chains of rules can be avoided.

Two cautions are in order, however. In interpreting a description (building the model in the first place), it suffices to place objects in simplest possible relationships. If a description mentions a book on a desk, we probably visualize the book as being directly on the desk. The reverse process--judging from a mental model whether a particular preposition is an appropriate description of the relation between two objects--is not always so simple. In deciding whether "above" is an acceptable description, for instance, there is little question when one object is directly above the other, but clearly the word is acceptable even when the direct case is not applicable, and deciding these more marginal cases often leads to a lot of hedging, even from native speakers.

The second caution is best put by describing a session with the implementation: as it happened, the particular mental model produced after

"a shelf is on a wall" and "a fly is on the wall" was the equivalent of the illustration below.



Now suppose we were to ask if the fly is under the shelf. The correct answer, of course, is "I don't know," since on the basis of the description the fly might be under the shelf, but it might be elsewhere, too. (If the implementation had been set up to try putting the fly under the shelf, and, subsequently, at a place violating the location restrictions of under, in response to the question, it would have found neither violated the location restrictions placed on the fly by the original description and hence would have had reason to suspect that it couldn't answer the question one way or the other.)

Clearly, then, the simple expedient of directly consulting the constructed model is a little too simple. The more freedom a model allows in choosing the location of an object, the more incidental any relations between various objects may be. In the end what we know are the location restrictions and it is based on them that we need to make judgments.

Regularities

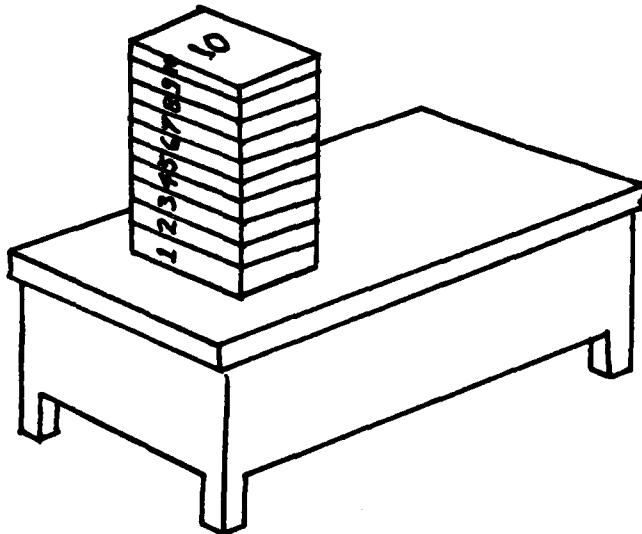
For all the hedges and caveats of the preceding paragraphs, it was evident from the implementation that paying attention to a very small set of attributes of objects yields an astonishing amount of descriptive power. The attributes included a very rudimentary surface description, the concept of a free-surface with associated free-direction, the essential "emptiness" of containers, some notion of gravity, of contiguity, of the interior relation in two or three dimensions, of partial axes of symmetry, some awareness of scale, and a coordinate system with marked vertical direction. Clearly, these concepts do not handle all cases of descriptions using place locatives. It might even be said that they do not handle some of the most common cases (we will come back to this in a moment). But they do handle the most typical cases--the regular uses of in, on, and the other prepositions--the uses we are most likely to think of as standard. In so doing, they capture much of the descriptive power of the prepositions.

Why then could it be said that they do not handle some of the most common cases? It is well known that the most frequently occurring verbs in English are also the most irregular. Something of the same sort seems to apply to uses of the prepositions with common objects. Tables, for instance, have a tendency to be treated as if they were essentially the table top--"under the table" for most objects means under the table top but definitely not under the legs. Rugs are an exception, of course, as are floors, and there are undoubtedly other exceptions to the mini-rule of treating the table as top only.

Is on transitive?

As another example of the irregularity of tables, consider a scene like that in figure 4, which

Figure 4



can be described by (12a - j):

(12a) Volume 10 is on volume 9.

(12b) Volume 9 is on volume 8.

⋮

(12i) Volume 2 is on volume 1.

(12j) Volume 1 is on the desk.

Since all volumes, 1 - 10 can be said to be "on the desk" we would like some kind of transitive rule to apply, but it would not be proper (or at least it would be very odd) to say that "Volume 10 is on volume 2." The hidden regularity here is that tables (and other furniture: desks, shelves, counters, etc.) have on relations with everything they support, directly or indirectly. Most other objects do not have on relations with everything they support, so that, for example the top book on a stack of books on the ground is not normally said to be "on the ground."

Fortunately, even the most common objects (including tables) appear to be regular most of the time, with most of the prepositions. It is

interesting that some of the irregularities fall into classes, like classes of irregular verbs (sing, sang, sung; drink, drank, drunk; sink, sank, sunk). For example, "the people on the bus" are actually in the bus--they aren't on the bus in the same sense that "the people on the car" would be on the car. On has the same interpretation in "on the plane," "on the subway," or "on the boat"--indeed for anything that can be boarded or alternatively that one can stand up in. So at least potentially there may be classes of irregular objects.

After all is said and done, though, it is still the case that the system seems to work, and work well, for the great majority of regular objects, and even for the irregular ones most of the time. It seems clear that basic understanding of the use of the prepositions is ours if only we pay attention to a small set of perceptually salient characteristics of the objects related.

6. Problems remaining

Clearly there will be surprises in programming the rest of the prepositions, and we have only begun to scratch the surface of the problems in implementing programs to deal with sentences like (1) - (3) (the "dog bites mailman" example). However, we are already aware of some problems and exceptions to the general picture presented in this paper.

One major problem was alluded to in the example in section 5 of the fly which was (arbitrarily) placed under a shelf in the mental model and could thereafter not be differentiated from a fly specifically asserted to be under the shelf. What seems to be needed is some way of keeping track of the range of possible positions available to objects described; we have debated several schemes (e.g. a probability distribution for

position, a tag on objects explicitly negating accidental relationships between objects, deferring the creation of a mental model until a question is raised, etc.) but are still undecided about the best way to proceed.

Another difficulty (initially pointed out to us by Phil Johnson-Laird) is that the preposition at seems to have the function of specifying a canonical relation between subject and object. Thus "the chair is at the desk" describes a specific relationship--if the chair is upside down or facing away from the desk, it can no longer be naturally said to be at the desk. Similarly at picks out canonical relations in "I stood at the window," "John was at the door," "I am at my desk," etc. At seems to require special scenarios for each object, and is otherwise regular only in that most scenarios require proximity of subject and object.

Many prepositions require that the positions of the speaker and/or listener with respect to the subject and object be known. For example, I could say to a listener in Japan that "Urbana is near Chicago," (it is about 120 miles away) but I would not say this to a listener 10 miles from Urbana (see also [Denofsky 1976]).

Most difficult (and most exciting) of the problems we are aware of are the transfers of meanings from the spatial domain to abstract domains. A representation of physical objects, events, and their relations should be able to be used in constructing effective representations for abstract phenomena. An important part of understanding the abstract use of prepositions involves identifying the "covert categories" to which words belong. As an example, consider the phrases below:

get { into a car
 { into trouble
 { into mischief }

be $\{ \begin{matrix} \text{in a car} \\ \text{in trouble} \\ \text{in mischief*} \end{matrix} \}$

get $\{ \begin{matrix} \text{out of a car} \\ \text{out of trouble} \\ \text{out of mischief*} \end{matrix} \}$

We suggest that both trouble and car belong to a covert category which could be called "spatial enclosures," but that mischief does not belong to this category, even though its meaning is much closer to trouble's than is car's meaning. This example seems to us to be similar to the mass/count distinction in English--words like house, person, and book are count nouns (we can say "a house" or "two houses") whereas sand, butter, and water are mass nouns (we cannot say "a sand" or "two sands," but must add a measure phrase, e.g. "a ton of sand," or "a lot of sand"). Mass nouns which common measures associated with them can sometimes be used as count nouns, as in "Waiter, bring me two waters," and some nouns, like paper, seem to fit equally well in either category. (Such categories are discussed in Whorf [1956].) We will not deal further here with transfer of meaning between domains, although this is a topic of great current interest to us. For those interested in this topic, Jackendoff [1975] is a fascinating source of ideas; also see Waltz [1978], and Pylyshyn [1977].

7. Related work

This research has been influenced by a number of other pieces of work. Several stand out and are described briefly in this section.

Three items stand out particularly: a thesis by N. Goguen [1973] a report by G. S. Cooper [1968], and a paper by H. H. Clark [1973]. Cooper's work developed a set of primitives and paper definitions for a number of prepositions. While the primitives proved to be inadequate when we began programming, this paper was an inspiration for the overall approach. Goguen

wrote a program in many ways similar to this, but did not address the problems of multiple interpretations of prepositions. Clark's paper provided valuable insights into the coordinate systems underlying spatial language, and into the types of mental models people create from scene descriptions.

D. V. McDermott's TOPLE [1974] deal with some very interesting aspects of building a "mental model" of a scene from natural language. For example, given the sentence.

(13) The banana is under the table, by the ball.

There are two interpretations: (1) the ball can be under the table, or (2) the ball can be near the table, but not under it. If we were given

(14) The banana is under the table, by the floor lamp.

then the interpretation where the floor lamp is near but not under the table becomes more likely, based on the typical size of a floor lamp. McDermott's program is able to use size to make this type of distinction. However, the "mental model" in this work is a data base of assertions, e.g.

(UNDER TABLE1 BANANA1)

and

(UNDER TABLE1 BALL1).

Winograd's SHRDLU [1972] is probably the most closely related program, though its tasks were rather different--its "mental model" was known completely to the language understander, not constructed by descriptive natural language input.

The book Language and Perception by Miller and Johnson-Laird [1976] is a valuable source of ideas and an excellent compendium of results from past work.

A number of other related publications are included in the list of references for the interested reader.

List of References

Black, F., "A Deductive Question-Answering System," in Minsky (ed.) Semantic Information Processing, Cambridge, MA: MIT Press, 1968.

Boggess, L. Computational interpretation of English Spatial prepositions. Ph.D. thesis, University of Illinois, 1978.

Clark, H. H. "Space, Time, Semantics, and the Child," in T. E. Moore (Ed.), Cognitive Development and the Acquisition of Language, New York: Academic Press, 1973.

Cooper, G. S. "A Semantic Analysis of English Locative Expressions." (Technical Report No. 1587), Cambridge, MA: Bolt, Baranek & Newman, Inc., 1968.

Cresswell, M. J. "Prepositions and Points of View," Linguistics and Philosophy, 1978, 2, 1-41.

Denofsky, M. E. "How Near Is Near? A Near Specialist." (AI Memo No. 344), MIT Artificial Intelligence Lab, 1976.

Fillmore, C. J. "The Case for Case Reopened," in J. P. Kimball (Ed.), Syntax and Semantics, New York: Academic Press, in press.

Goguen, N. A procedural description of spatial prepositions. M.S. thesis, Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, 1973.

Jackendoff, R. "A System of Semantic Primitives," in Schank and Nash-Webber (eds.), Theoretical Issues in Natural Language Processing, ACL, Arlington, Virginia, 1975.

Johnson-Laird, P. "Mental Models of Meaning." Paper presented at Sloan Workshop on Computational Aspects of Linguistic Structure and Discourse Setting, University of Pennsylvania, May 1978. To appear in a book A. Joshi (ed.) Cambridge University Press, 1980.

Kosslyn, S. M. and Schwartz, S. P., "A Simulation of Visual Imagery," Cognitive Science 1, 3, July 1977, 265-95.

McDermott, D. Assimilation of New Information by a Natural Language-Understanding System. Cambridge, MA: MIT AI Lab TR241, 1974.

Miller, G. A. and Johnson-Laird, P. Language and Perception, Cambridge, MA: Harvard University Press, 1976.

Piaget, J. and Inhelder, B. The Child's Conception of Space, translated by F. J. Langdon and J. L. Lunzer, New York: W. W. Norton & Co., 1967.

Pratt, V. R. "A Linguistics-Oriented Programming Language," Memo No. 277, MIT AI Lab, Cambridge, MA, 1973.

Pylyshyn, Z. W. "Children's Internal Description," in Language, Learning, and Thought. New York: Academic Press, 1977.

Schank, R. C. and Abelson, R. P. Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

Sondheimer, N. K. 'Spatial Reference and Semantic Nets,' American Journal of Computational Linguistics, 1977, 71, 1-67.

Sussman, G. J. "The FINDSPACE Problem," Memo No. 286, MIT AI Lab, Cambridge, MA, 1973.

Waltz, D. L. "On the Interdependence of Language and Perception," in TINLAP-2: Theoretical Issues in Natural Language Processing-2, ACM, New York, 1978.

Whorf, B. L. "A Linguistic Consideration of Thinking in Primitive Communities," in J. B. Carroll (Ed.), Language, Thought and Reality, Cambridge, MA: The MIT Press, 1956.

Wilks, Y. Grammar, Meaning and the Machine Analysis of Language, London and Boston: Routledge, 1972.

Winograd, T. Understanding Natural Language, New York: Academic Press, 1972.